지능형 설계자동화 2020년 1학기

Class Activation Mapping(CAM)

SMART DESIGN LAB
유 소 영

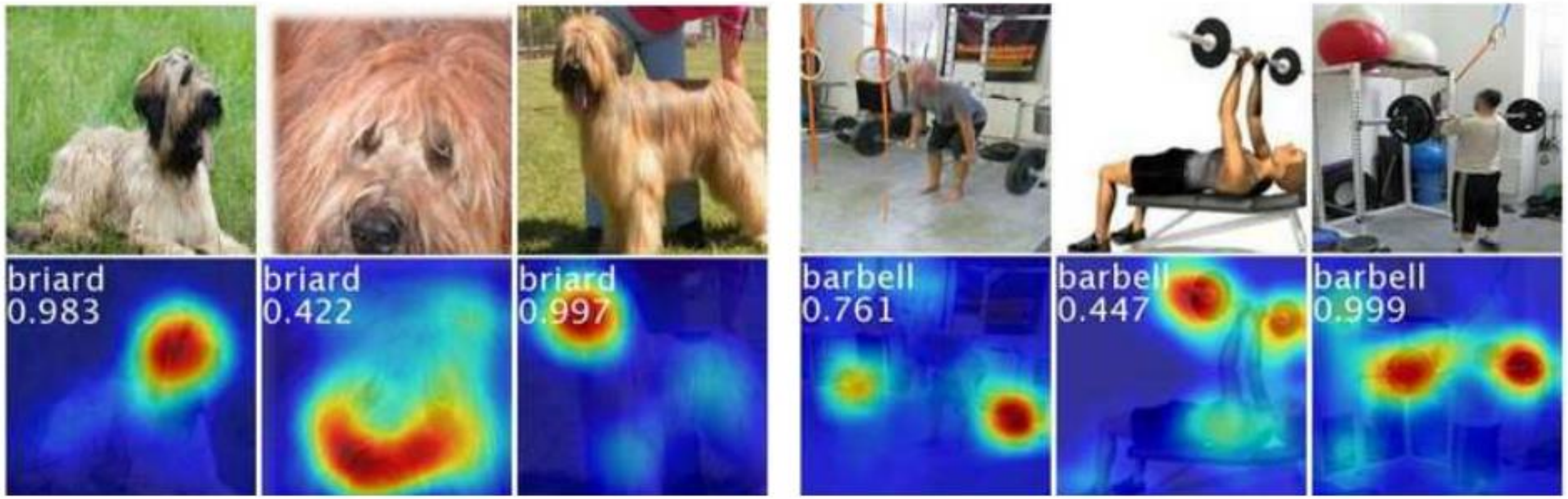
2020년 4월 22일

Contents

Introduction

CAM의 핵심은 FC 대신에 GAP을 쓰는것.

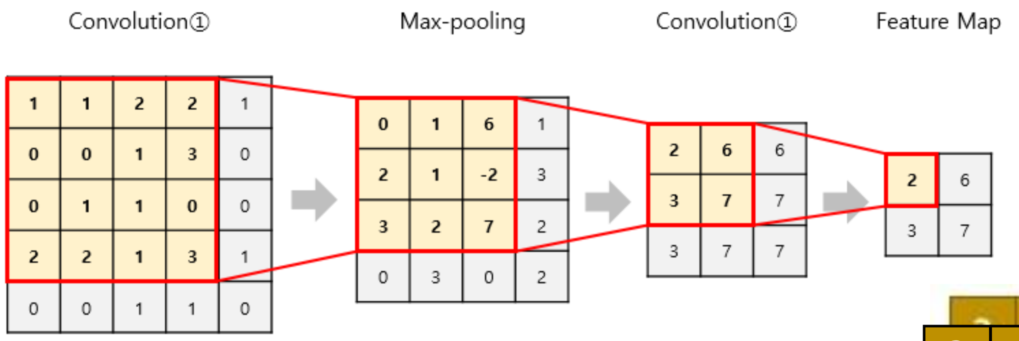
- B. Zhou, A. Khosla et al., “Learning Deep Features for Discriminative Localization,” CVPR 2016
- Zhou et al has shown that convolutional neural networks (CNNs) behave as object detectors
- This ability is lost when **fully-connected** layers are used for classification
- In the experiments, A **global average pooling(GAP)** layer is shown advantages beyond simply acting as a **regularizer**
- a little tweaking, the network can retain its remarkable localization ability until the final layer.



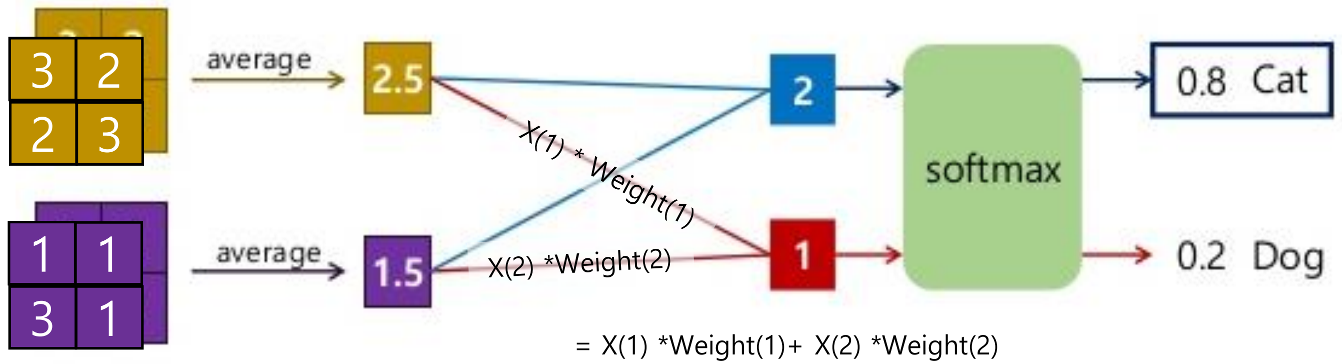
https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Zhou_Learning_Deep_Features_CVPR_2016_paper.html

Proposed method : Global Average Pooling(GAP)

- M. Lin, Q. Chen, and S. Yan, "**Network In Network**," ICLR 2014.
- In the last convolution layer of CNNs
  - Replace the traditional fully connected layers with global average pooling
    - Generate **one feature map** for each corresponding category
    - Take the average of each feature map, and the resulting vector is fed directly in to the softmax layer



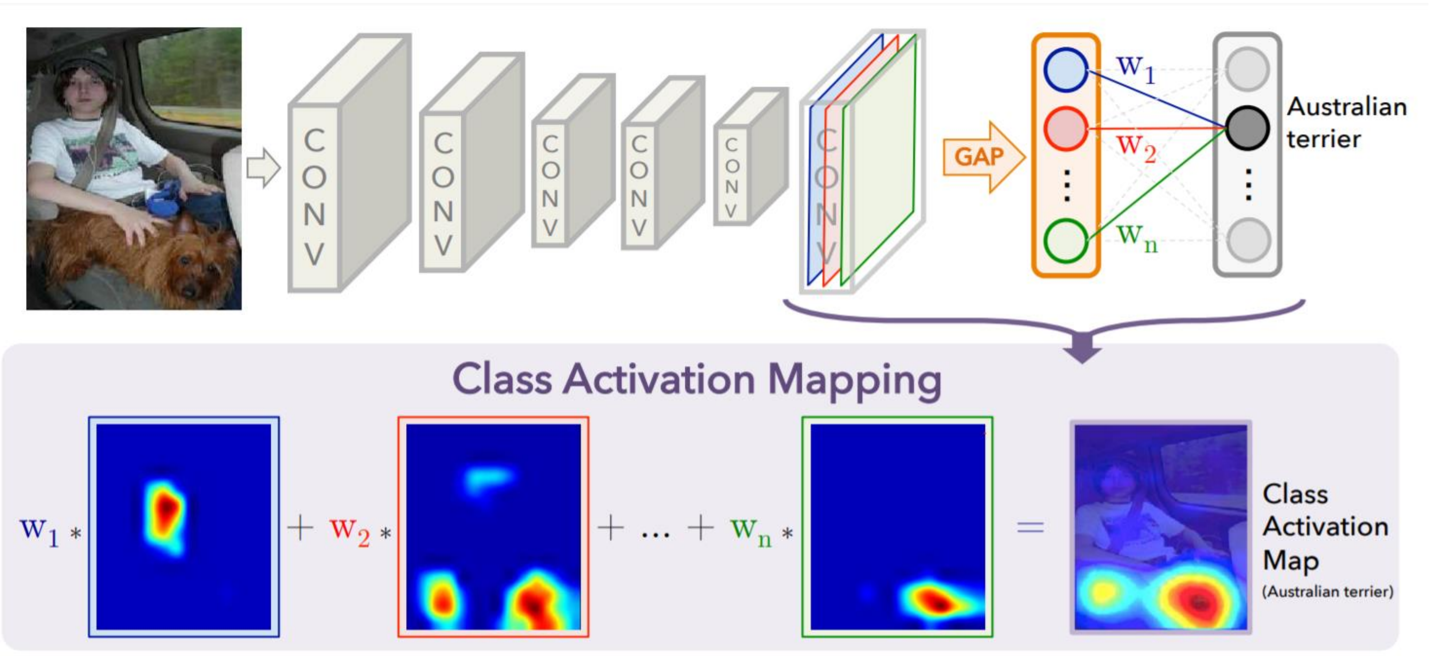
연산이 필요한 **파라미터 수를 크게 줄이는** regularizer 역할로써 **과적합을** 방지함.



* 유의! GAP은 원래 전체 평균을 내는 것 이지만 본 논문에서는 전체 합을 취한 것 으로 GAP을 표현했음

Proposed method : Class Activation Mapping (1)

- CAM(Class Activation Mapping) represent discriminative image regions used by the CNN

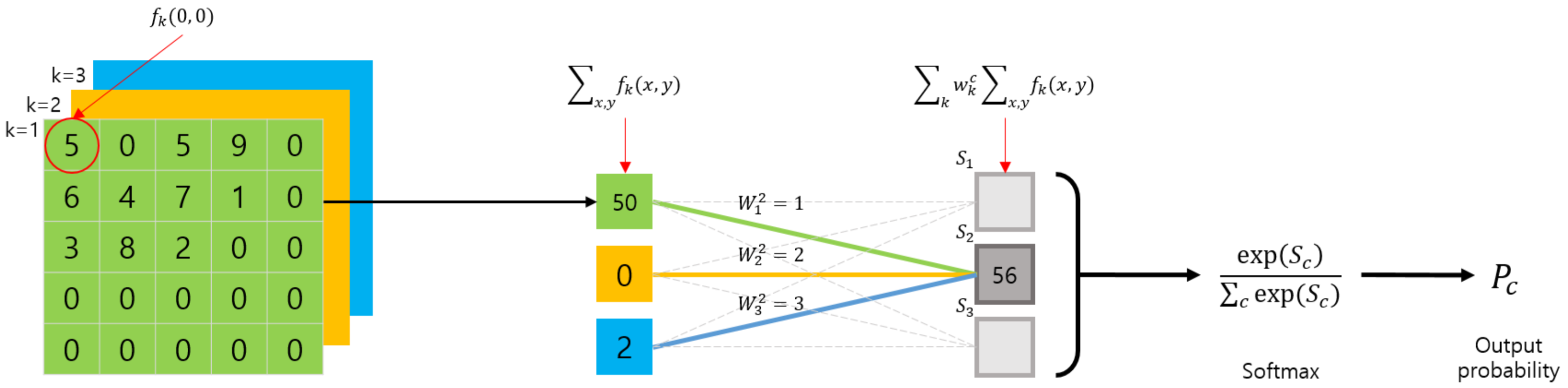


- (1) Feature map(14 X 14 X N)을 Global Average Pooling(GAP)함
- (2) GAP을 통해 길이 N인 vector 가 출력됨
- (3) 가장 높은 확률을 가지는 Class를 예측하면 각 GAP output인 vector들의 weight을 뽑아 낼 수 있음
- (4) weight를 Feature map에 weighted sum하게 되면 위의 이미지와 같은 Heat map 생성됨
- (5) 이 Heat map을 원본 이미지 크기로 resizing 하고 Overlay하면 Class Activation maps(CAM)이 생성됨

Proposed method : Class Activation Mapping (2)

*사실, GAP은 평균을 취해야 하지만 논문에서 표현한 수식으로 설명함.

- CAM Architecture
  - $f_k(x,y)$: k번째 feature map



$f_k(x,y)$

Last convolution layer output

$F_k = \sum_{x,y} f_k(x,y)$

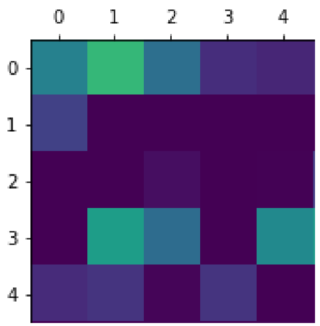
GAP output

$$S_c = \sum_k w_k^c F_k = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y)$$

- CAM
  - S_c : Class c의 예측 Score
  - M_c : CAM
  - M_c(x,y) $=\sum_k w_k^c f_k(x,y)$

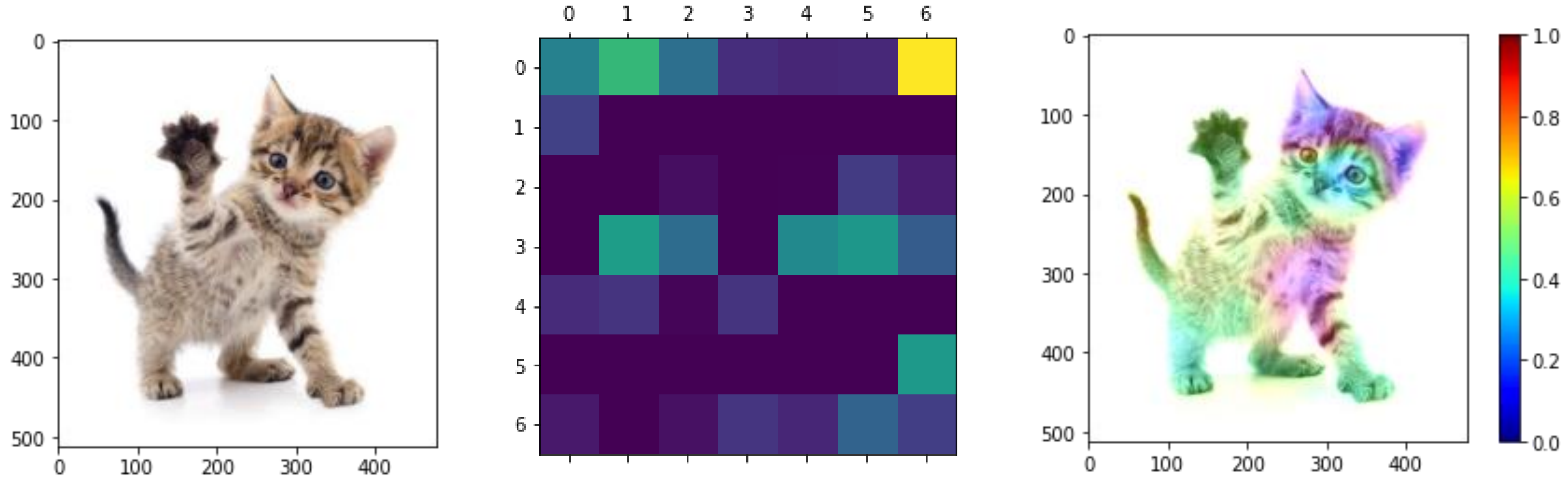
$$\begin{bmatrix} 5 & 0 & 5 & 9 & 0 \\ 6 & 4 & 7 & 1 & 0 \\ 3 & 8 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \boldsymbol{w_1^2} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \boldsymbol{w_2^2} + \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \boldsymbol{w_2^3} =$$

CAM



Tutorial

CAM Source Code:
https://drive.google.com/open?id=1tulxGCNMYYm4pipR9xSskrED4kUkoIHC

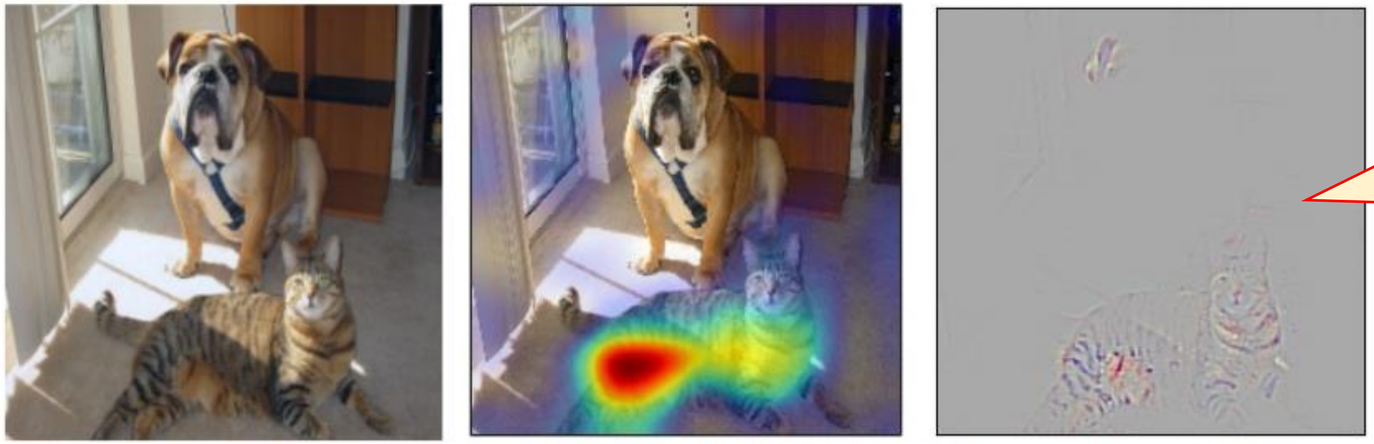


Cons

- CAM의 가장 큰 단점은 Global Average Pooling layer가 반드시 필요하다는 점.
  - GAP이 이미 포함되어 있는 GoogLeNet의 경우에는 문제가 없겠지만,
  - 그렇지 않은 경우에는 마지막 convolutional layer 뒤에 GAP를 붙여서 다시 fine-tuning 해야 함
  - 약간의 성능 감소를 동반하는 문제가 있음.
- 마지막 layer에 대해서만 CAM 결과를 얻을 수 있는 점
- CAM은 image 상에서 class와 관련된 부분을 대략적으로는 잘 찾아내지만, Upsampling영향으로 그 부분의 detail은 잘 잡아내지 못함.

[관련논문] Gradient based CAM

- RR Selvaraju, M Cogswell et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization" ICCV 2017
- Our approach uses the gradients for highlighting important regions in the image
- Grad-CAM is applicable to a wide variety of CNN model:
  - (1) CNNs with fully-connected layers
  - (2) CNNs used for structured outputs
  - (3) CNNs used in tasks with multi-inputs or reinforcement learning, without any architectural changes or re-training.



(a) Original Image (c) Grad-CAM 'Cat' (d)Guided Grad-CAM 'Cat'

또한, 저자는 Grad-CAM과 Guided Backpropagation의 결과를 element-wise multiplication을 해서 얻을 수 있는 **Guided Grad-CAM**을 제안.

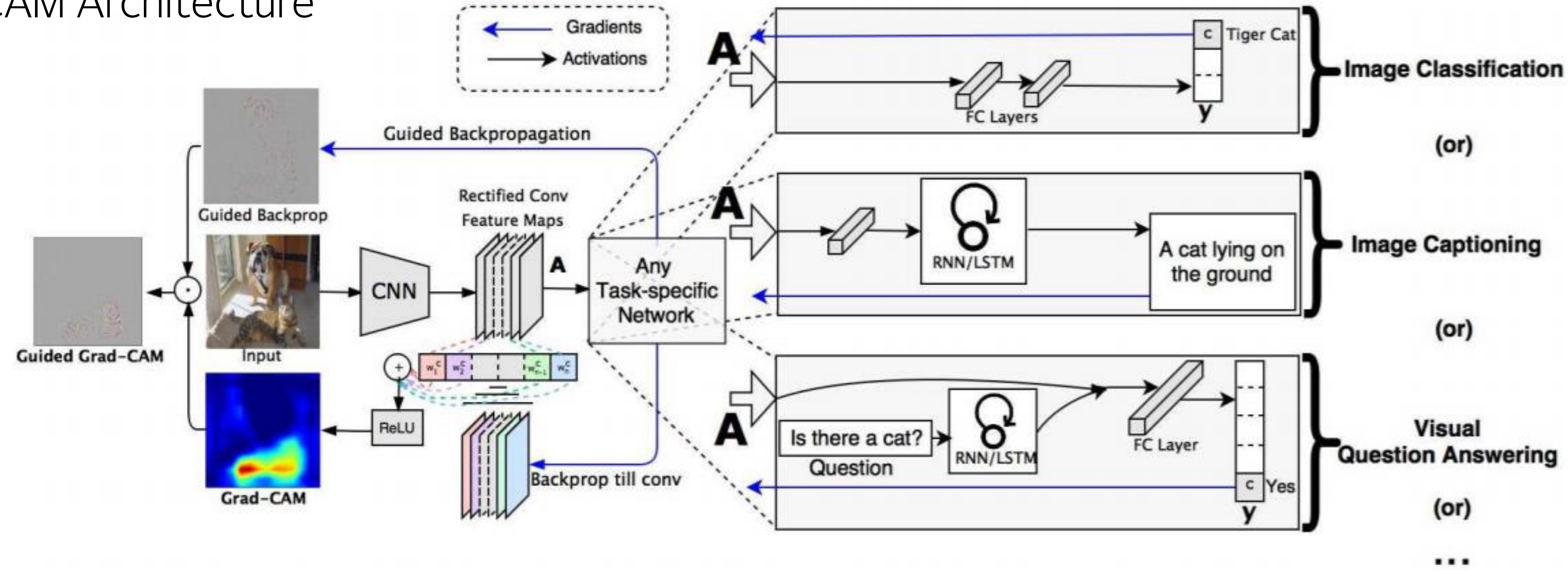
$$\begin{array}{c} G \\ \begin{bmatrix} 3 & 5 & 7 \\ 4 & 9 & 8 \end{bmatrix} \end{array} \circ \begin{array}{c} H \\ \begin{bmatrix} 1 & 6 & 3 \\ 0 & 2 & 9 \end{bmatrix} \end{array} = \begin{array}{c} N \\ \begin{bmatrix} 3\times 1 & 5\times 6 & 7\times 3 \\ 4\times 0 & 9\times 2 & 8\times 9 \end{bmatrix} \end{array}$$

element-wise multiplication

https://arxiv.org/abs/1610.02391

[관련논문] Gradient based CAM (2)

- Grad CAM Architecture



Softmax 입력 y^c가 feature map $A^k_{i,j}$에 대해 가지는 gradient

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

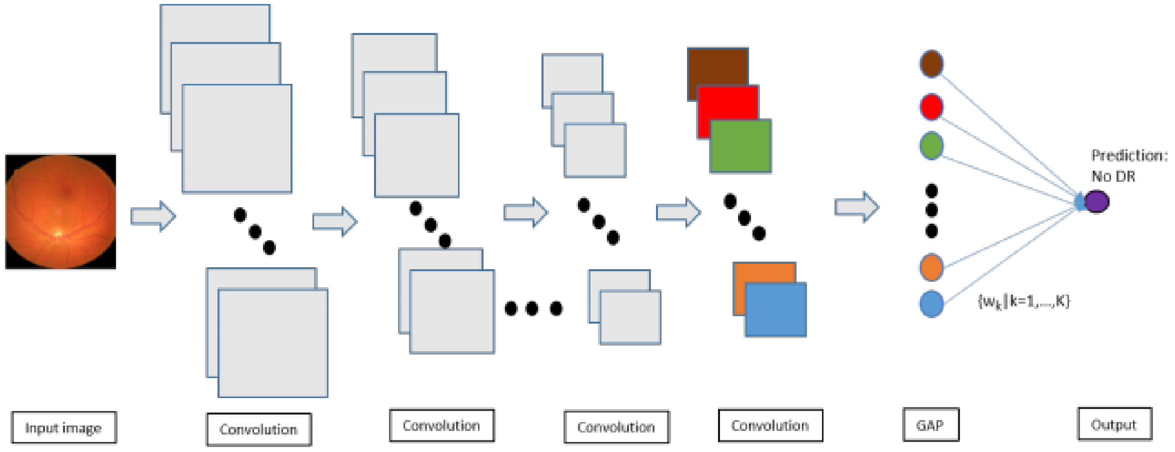
gradient를 GAP으로 표현 한 값

$$L^c_{\text{Grad-CAM}} = ReLU\underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}}$$

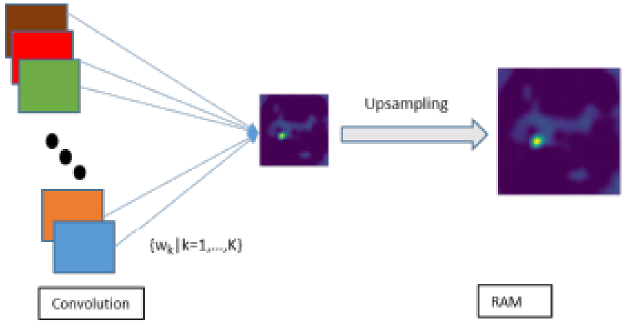
즉, CAM과 달리 아키텍쳐 변경이나 재학습을 하지 않고도 Grad CAM을 구할 수 있다.

[관련논문] Regression Activation Mapping

- RAM was Inspired by CAM
- Localize the discriminative region towards the **regression outcomes**.
- Using GAP and the linear output unit

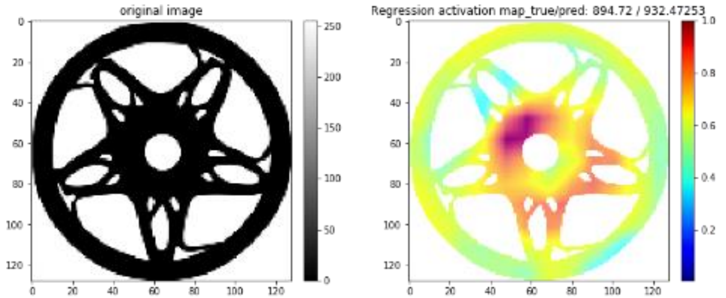


(a)



(b)

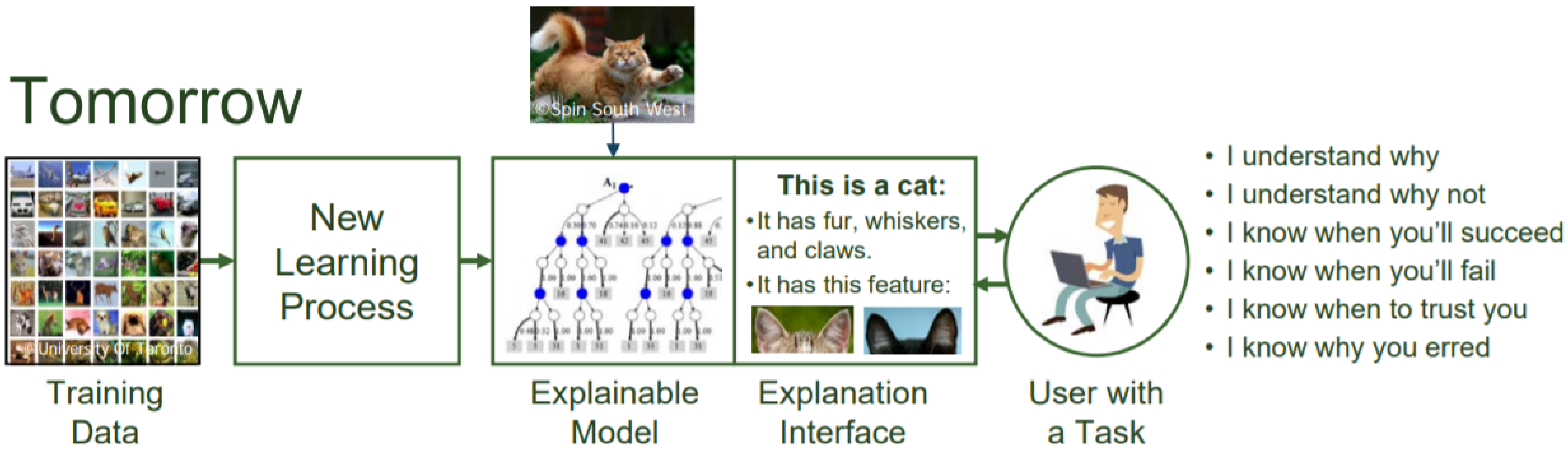
- **Example) Wheel mode frequency regression**



https://arxi

Conclusion

- CAM , Grad CAM , RAM 을 통해 Explainable Artificial Intelligence (XAI)을 구현할 수 있음.
- 딥러닝 모델 구현에 대한 근본적인 이해가 가능하여
- 모델에 대한 원인을 파악해 AI 모델의 성능 향상에 도움이 됨
- AI 모델에 대한 이해를 바탕으로 잘 학습된 모델을 설명할 수 있게됨.



- Thanks!

Reference

- https://www.darpa.mil/attachments/XAIProgramUpdate.pdf
- https://github.com/zhoubolei/CAM
- https://www.slideshare.net/healess/explainable-ai-xai-167954957
- "Explainable Artificial Intelligence (XAI)". DARPA presentation. DARPA. Retrieved 17 July 2017.
- https://jetsonaicar.tistory.com/16
- https://you359.github.io/cnn%20visualization/GradCAM/